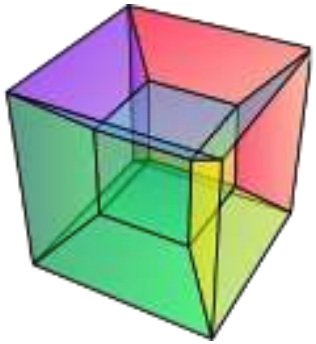
HyperSpy Community & Github

Eric Prestat

University of Manchester and SuperSTEM Laboratory



- HyperSpy on the web
- What is Github? Why is it so popular?
- What makes the HyperSpy community?

'Enhanced Data Generated by Electrons'

Following the very successful workshops at Lake Tahoe, Leukerbad, Port Ludlow, Guadeloupe, Grundlsee and Banff, the next international workshop on electron energy loss spectroscopy and imaging will be held in

Sainte Maxime, France,

- Francisco's talk on HyperSpy
  - Mostly about open source and scientific python community
  - NOT about the cool stuff that HyperSpy could already do at the time

- Open source is not only about opening the source code, it is also about:
  - Accessibility to the users
  - Sustainable library development
  - Building a community

The University of Manchester

HyperSpy
multi-dimensional data analysis

HyperSpy: multi-dimensional data analysis toolbox

HyperSpy is an open source Python library which provides tools to facilitate the interactive data analysis of multi-dimensional datasets that can be described as multi-dimensional arrays of a given signal (e.g. a 2D array of spectra a.k.a spectrum image).

HyperSpy aims at making it easy and natural to apply analytical procedures that operate on an individual signal to multi-dimensional arrays, as well as providing easy access to analytical tools that exploit the multi-dimensionality of the dataset.

Its modular structure makes it easy to add features to analyze different kinds of signals.

Highlights

- Two families of named and scaled axes: *signal* and *navigation*.
- Visualization tools for multi-dimensional spectra and images.
- Easy access multi-dimensional curve fitting and blind source separation.
- Built on top of NumPy, SciPy, matplotlib and scikit-learn.
- Modular design for easy extensibility.

SuperSTEM DARESBURY

https://hyperspy.org

HyperSpy documentation/community support

The University of Manchester

- Main website:
  - https://hyperspy.org/
- Documentation:
  - https://hyperspy.org/hyperspy-doc/current/index.html
- Gitter: online chat room
  - https://gitter.im/hyperspy/hyperspy
- Users list:
  - https://groups.google.com/forum/#!forum/hyperspy-users
  - Not used much anymore in favour of gitter

SuperSTEM DARESBURY

HyperSpy demos

The University of Manchester

- Source code of the demo on github
  - https://github.com/hyperspy/hyperspy-demos
- Non-interactive version on nbviewer
  - https://nbviewer.jupyter.org/github/hyperspy/hyperspy-demos/tree/master/
  - Github now displays nicely the notebook, but works well for short notebook (loading fail otherwise)
- Interactive version on mybinder.org
  - https://mybinder.org/v2/gh/hyperspy/hyperspy-demos/master
  - Run demos online (on a remote server) without any installation
  - 1 CPU core, between 1 and 2 GB available memory
  - Shut down after 10 min inactivity

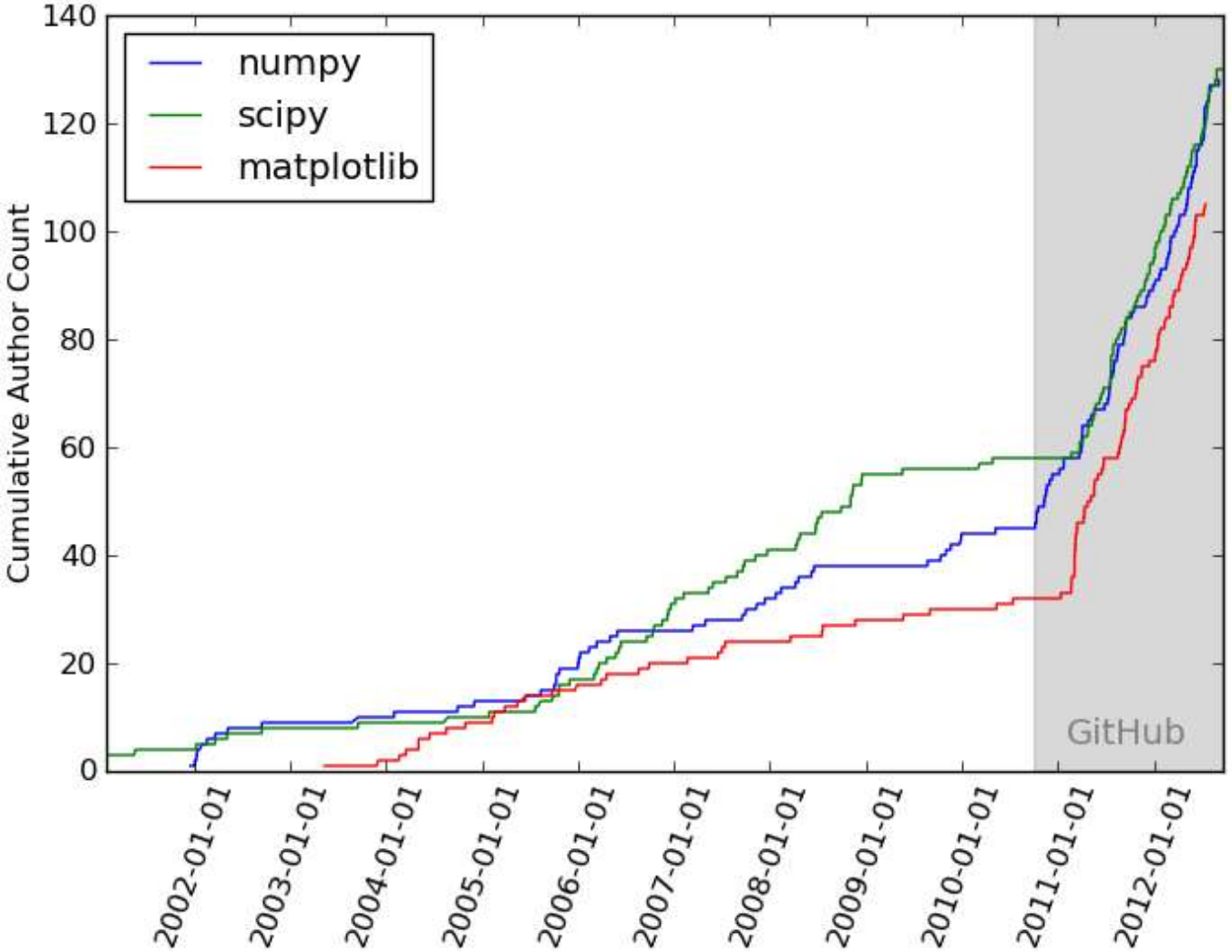
HyperSpy development

The University of Manchester

- Development site on Github
  - https://github.com/hyperspy/hyperspy
  - Issues tracker: report bug, propose new features and other business
  - Pull requests: discuss the merge of any changes with the upstream branches
- Developer guide
  - http://hyperspy.readthedocs.io/en/stable/dev_guide.html

The github effect

The University of Manchester



From Pythonic perambulations: Why Python is the last language you'll have to learn:
https://jakevdp.github.io/blog/2012/09/20/why-python-is-the-last/

SuperSTEM DARESBURY

What is github?

The University of Manchester

- At the heart of Github is git
  - A version control system: manages and stores revisions of projects
  - git is a distributed version control systems
    - Each contributor has is own *remote* (online) repository
    - Code is merged in the *upstream* repository

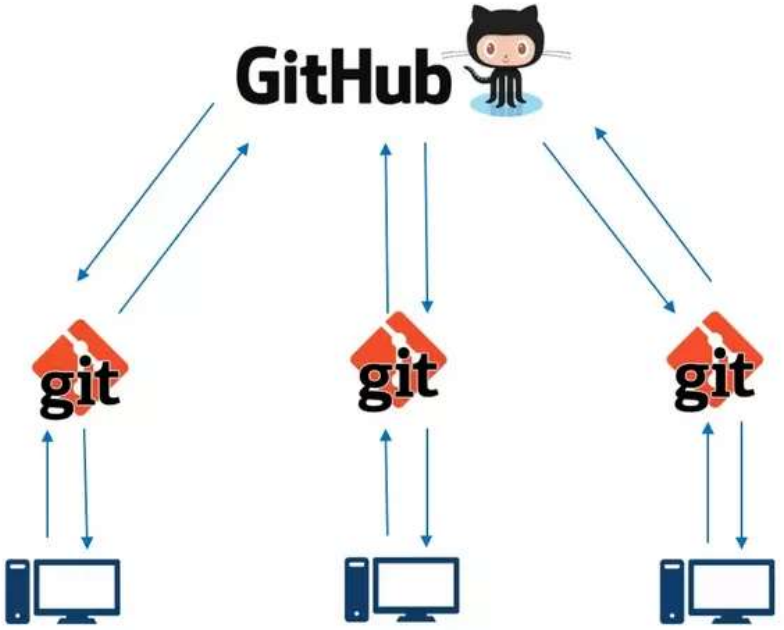
SuperSTEM DARESBURY

Short video explaining github: https://www.youtube.com/watch?v=w3jLJU7DT5E

The University of Manchester

What is github?

- GitHub is a git repository hosting service with its own features
  - Web-based graphical interface
  - provides a centralised place where people discuss the changes
  - “Social coding”
    - Open discussion in the issue tracker or PR submission, wiki, etc

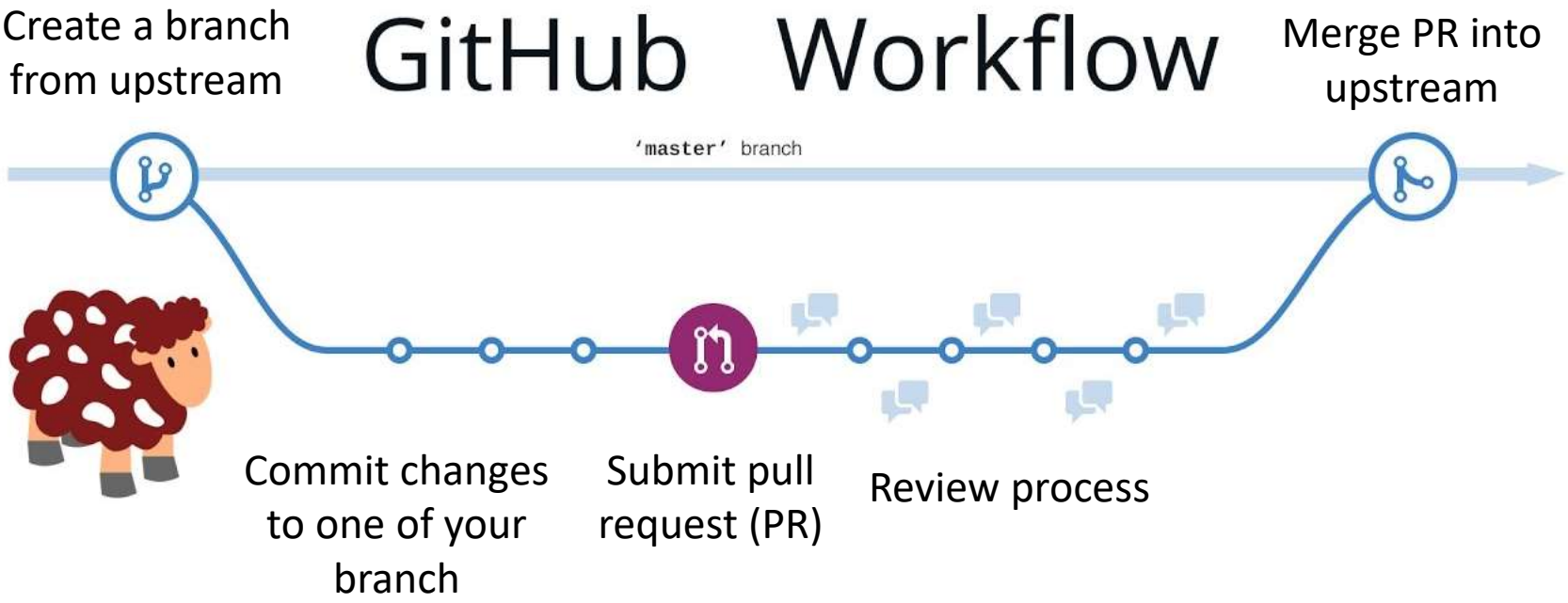


SuperSTEM DARESBURY

Short video explaining github: https://www.youtube.com/watch?v=w3jLJU7DT5E

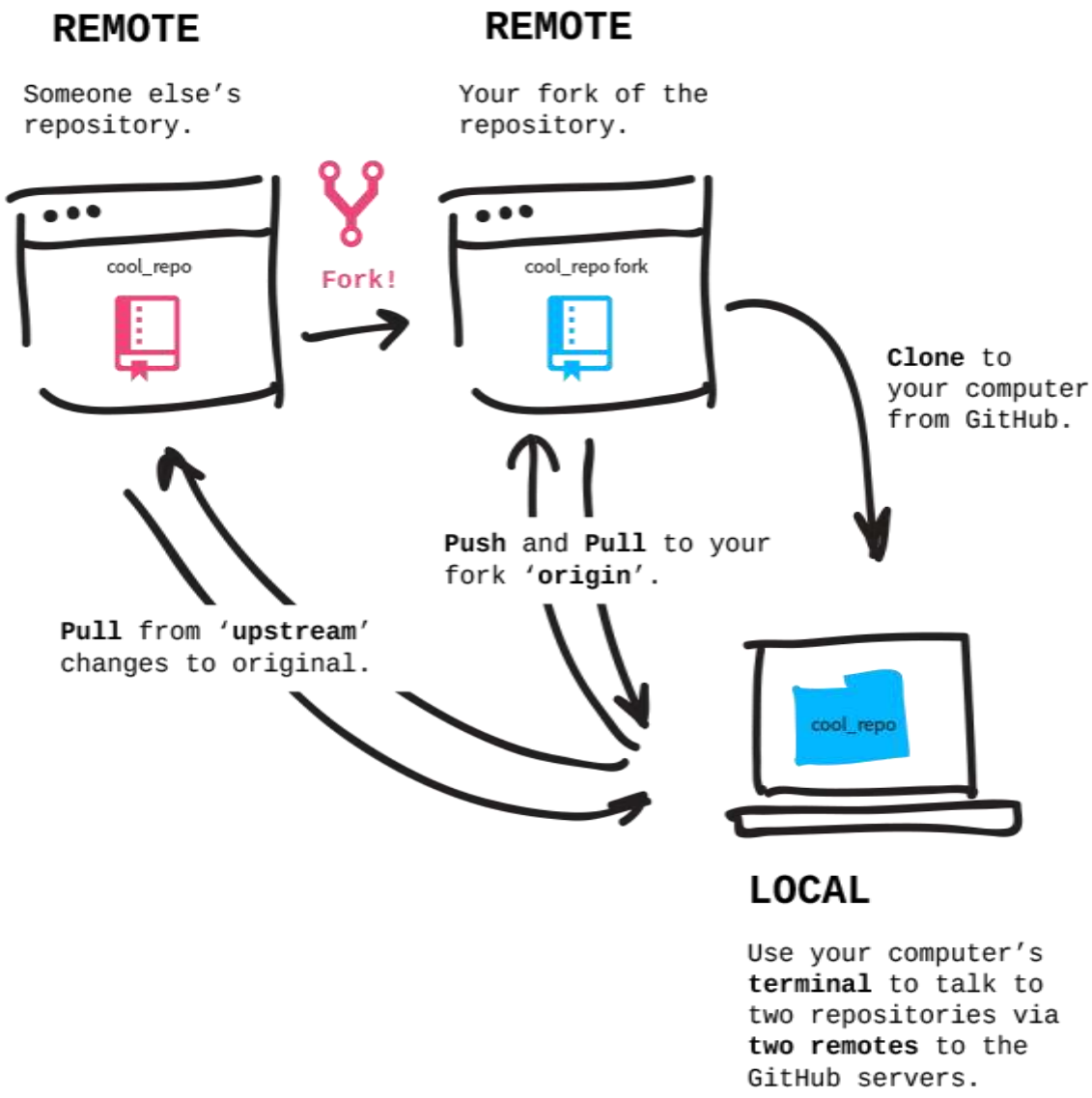
How to use github?

The University of Manchester



SuperSTEM DARESBURY

How to use github?



Contribute to an open-source project before git/github

The University of Manchester

1. manually download the project's source code
2. make your changes locally
3. create a list of changes called a 'patch' and then e-mail the patch to the project's maintainer
4. The maintainer would then have to evaluate this patch, possibly sent by a total stranger and decide whether to merge the changes.

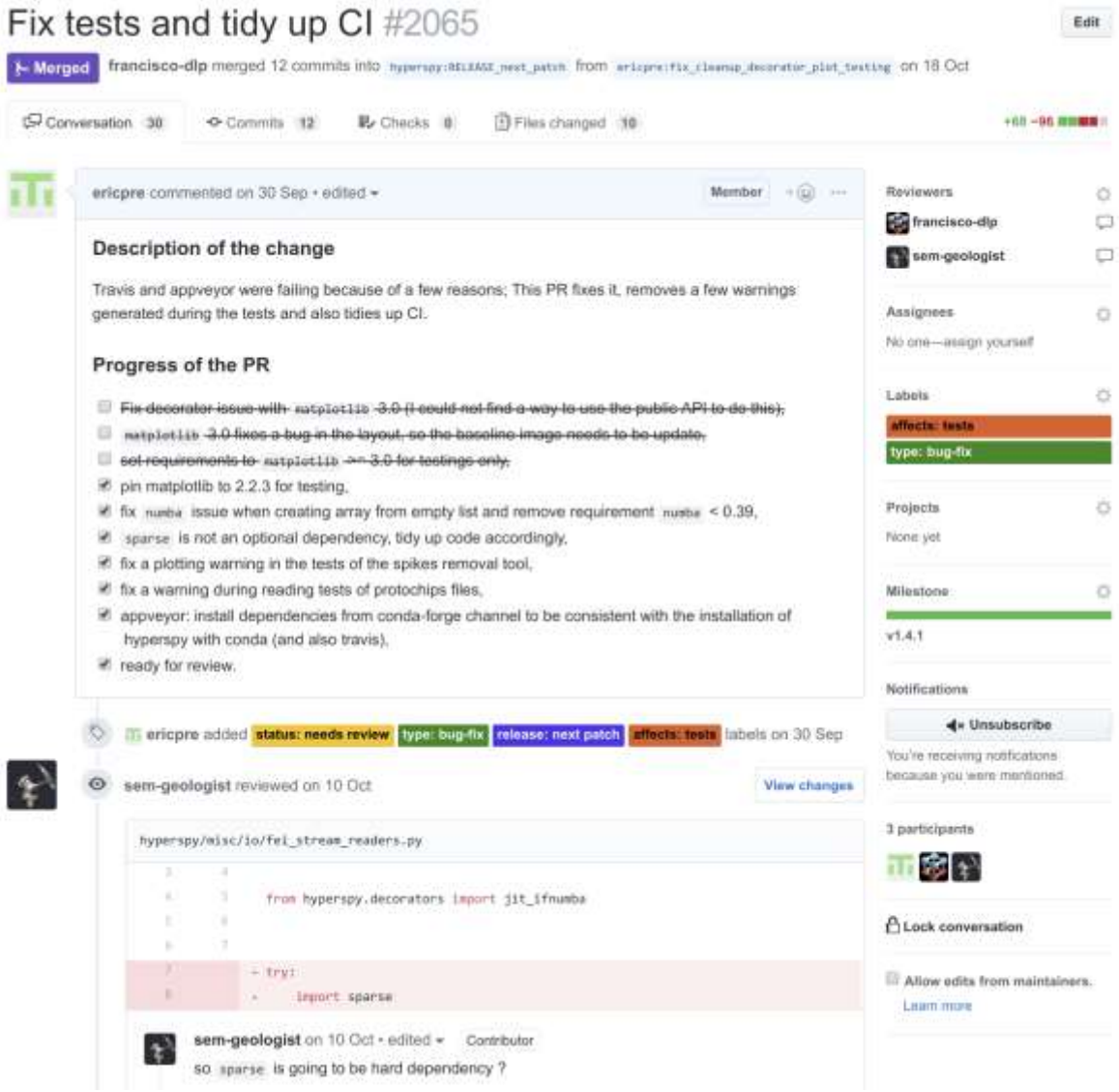
Tedious and not manageable

SuperSTEM DARESBURY

Why is github so useful/popular?

github web interface makes things easier

- Share code with other developers and the public
- Manage issues
- Manage pull request (code comparison and review)
- Maintain code



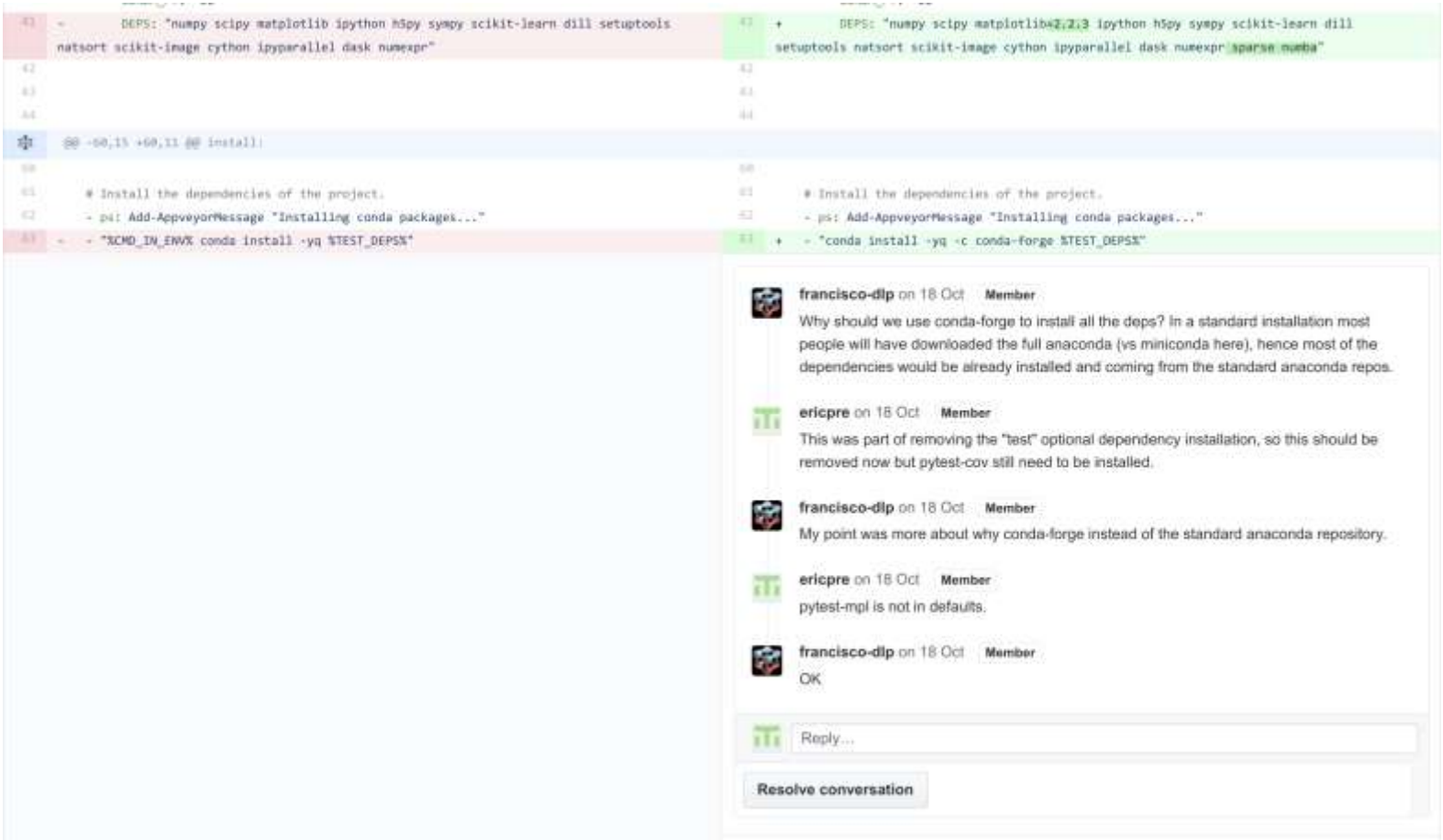
The University of Manchester

SuperSTEM DARESBURY

Why is github so useful/popular?

github web interface makes things easier

- Review and discuss changes during pull request (PR) review



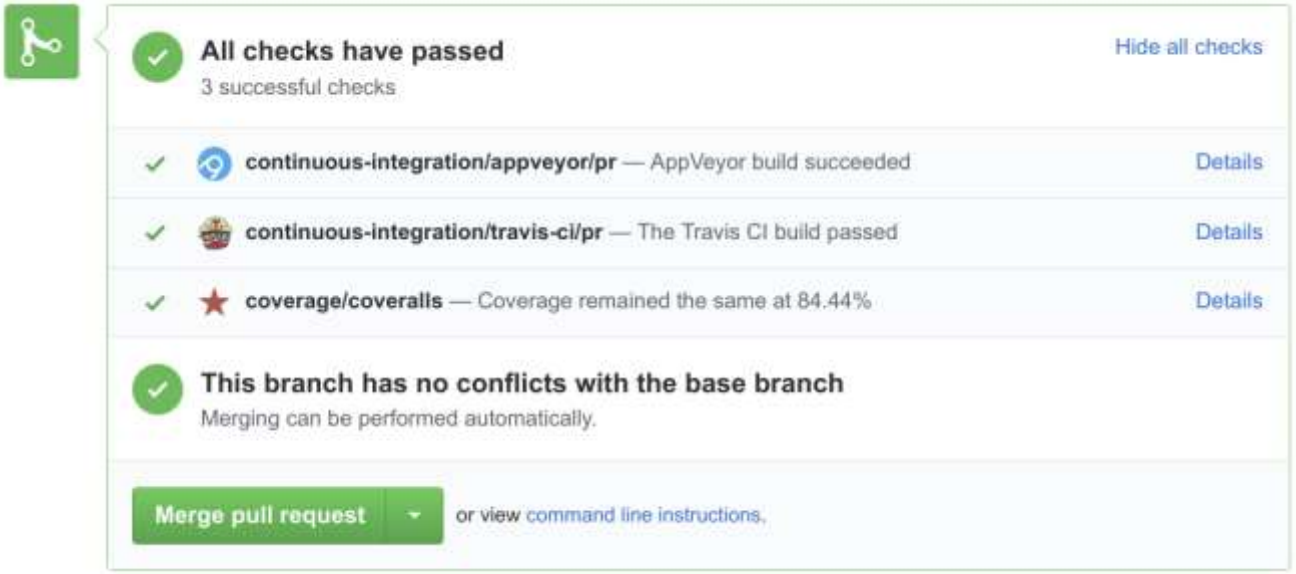
The University of Manchester

SuperSTEM DARESBURY

Why is github so useful/popular?

The University of Manchester

- Continuous integration
  - For each PR, the code is tested automatically against a suite of tests using external services (Github Actions, Azure, etc.)
  - ~4220 *unit tests* continuously checking that no regression is introduced by new changes

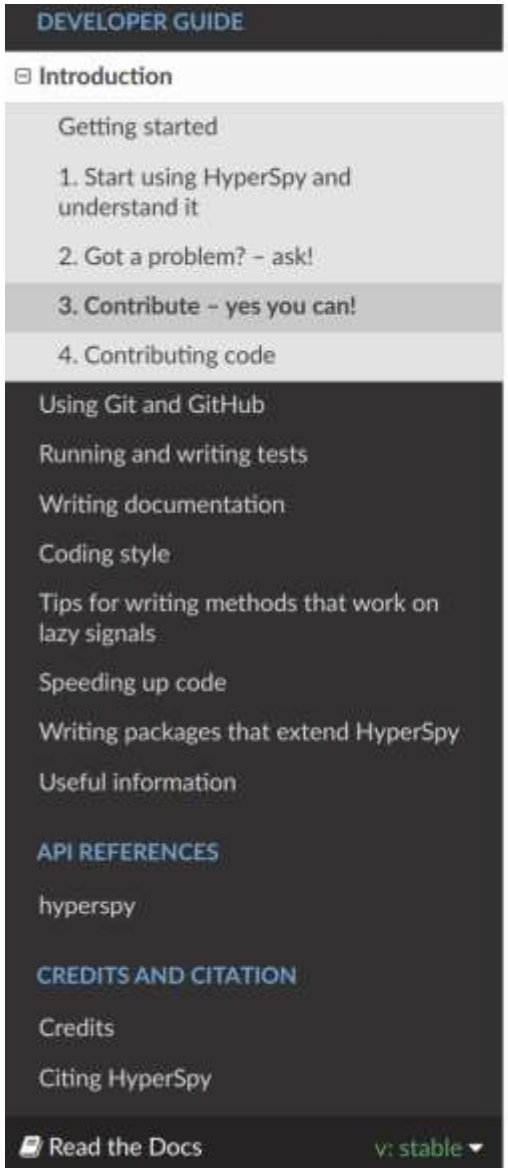


SuperSTEM DARESBURY

https://en.wikipedia.org/wiki/Unit_testing

HyperSpy developer guide

The University of Manchester



3. Contribute – yes you can!

You don't need to be a professional programmer to contribute to HyperSpy. Indeed, there are many ways to contribute:

1. Just by asking a question in our Gitter chat room instead of sending a private email to the developers you are contributing to HyperSpy. Once you get more familiar with HyperSpy, it will be awesome if you could help others with their questions.
2. Issues reported in the issues tracker are precious contributions.
3. Pull request reviews are essential for the sustainability of open development software projects and HyperSpy is no exception. Therefore, reviews are highly appreciated. While you may need a good familiarity with the HyperSpy code base to review complex contributions, you can start by reviewing simpler ones such as documentation contributions or simple bug fixes.
4. Last but not least, you can contribute code in the form of documentation, bug fixes, enhancements or new features. That is the main topic of the rest of this guide.

4. Contributing code

You may have a very clear idea of what you want to contribute, but if you're not sure where to start, you can always look through the issues and pull requests on the GitHub Page. You'll find that there are many known areas for development in the issues and a number of pull-requests are partially finished projects just sitting there waiting for a keen new contributor to come and learn by finishing.

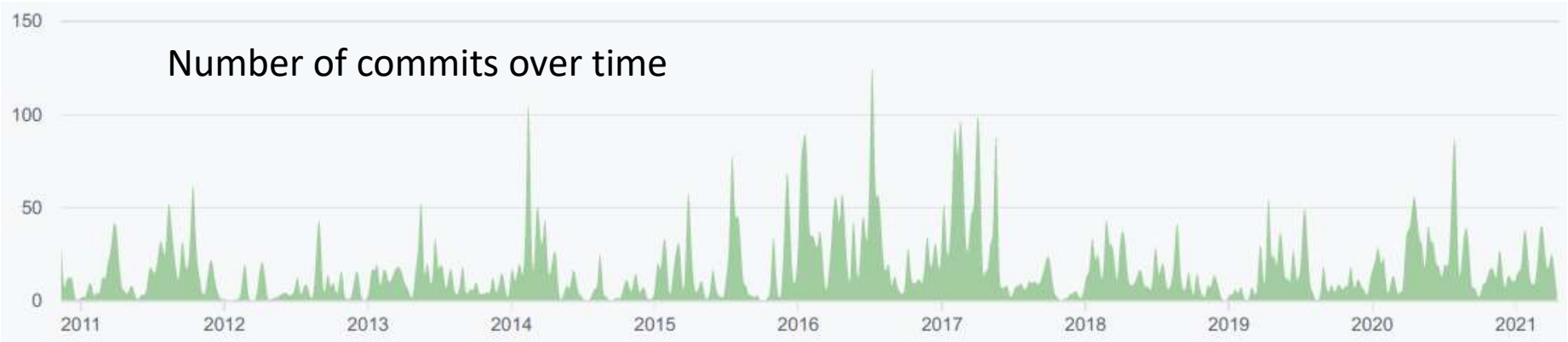
SuperSTEM DARESBURY

https://hyperspy.readthedocs.io/en/stable/dev_guide/intro.html

HyperSpy community

The University of Manchester

- Code contributors
  - ~49 contributors in total from many different labs
  - A few contributors change jobs
    - Their github profile may have been useful for their successful application
  - ~ 20 of them are a one off contribution



- Many more people following the gitter chat
  - Bug report, user feedback, requests, etc.

SuperSTEM DARESBURY

The HyperSpy eco-system last year

- Extension registration: other libraries can create their specific Signal

The University of Manchester

HyperSpy-ipywidget

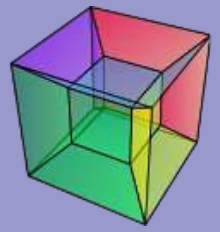
HyperSpy-traitsui

HyperSpyUI

EELS

Machine learning

Big data



EDS

Multi-dimensional

IO

Visualisation and interactivity

holography

Curve-fitting

Atomic position fitting

Quantitative STEM

Atomap

Strain mapping

4D STEM

Magnetic measurement

pixStem

Differential contrast imaging

Strain mapping

Orientation mapping

pyXem

Electron cristallography

Scanning Electron diffraction

cathodoluminescence

LumiSpy

photoluminescence

EBSD

Strain mapping

Kikuchipy

Orientation mapping

Electron Correlation Microscopy

Angular Correlations

EMpyer

Fluctuation Electron Microscopy

inpainting

inpystem

reconstruction techniques

segmentation

ParticleSpy

Particles analysis

Other extension to come

hyperspy_swift_library

Access nionswift libraries

SuperSTEM DARESBURY

The HyperSpy eco-system as of today

- Extension registration: other libraries can create their specific Signal

The University of Manchester

HyperSpy-ipywidget

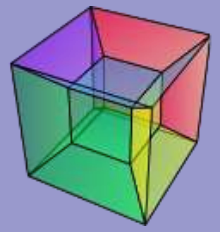
HyperSpy-traitsui

HyperSpyUI

EELS

Machine learning

Big data



EDS

Multi-dimensional

IO

Visualisation and interactivity

holography

Curve-fitting

Atomic position fitting

Quantitative STEM

Atomap

Strain mapping

Differential contrast imaging

Electron cristallography

pyXem

Scanning Electron diffraction

Pair distribution function

4D STEM

cathodoluminescence

LumiSpy

photoluminescence

EBSD

Strain mapping

Kikuchipy

Orientation mapping

Other extension to come

inpainting

inpystem

reconstruction techniques

segmentation

ParticleSpy

Particles analysis

pyMatchSeries

Non-rigid image registration

hyperspy_swift_library

Access nionswift libraries

pgure-svt

Denoising time-resolved microscopy

SuperSTEM DARESBURY

Future HyperSpy ecosystem and other relevant libraries

The University of Manchester

EELS

IO

EDS

Holography

HyperSpyUI

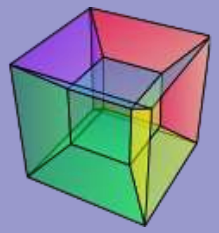
HyperSpy-ipywidget

HyperSpy-traitsui

Multi-dimensional

Curve-fitting

Visualisation and interactivity



Big data

Machine learning

pycroscopy

GMS 3

LiberTEM

tomviz

Nionswift

HyperSpy extension

HyperSpy aware library

SuperSTEM DARESBURY

Python libraries for 4D STEM data analysis

The University of Manchester

- pyXem
- pixstem
- EMpyer
- py4DSTEM
- pycroscopy
- stemtools
- Various nionswift plugins
- LiberTEM
- …

All these open source libraries are doing very similarly things

Optimised for high throughput on small computer clusters

SuperSTEM DARESBURY

Common issues with research software development

The University of Manchester

- Many publications present interesting data analysis workflow or methodological development
- These are generally not easily accessible
  - proprietary licence
  - not easily to install or to run (platform dependent, etc.)
  - not supported/up to date anymore
- May be difficult to maintain for the developer
- Implementation need to be generic and robust
- Rely on a specific research group (most of the time by one or a few PhD student or a post-doc..)
  - Not sustainable

SuperSTEM DARESBURY

Package distribution

The University of Manchester

- Integration with other libraries and workflow
- Installation needs to be easy

- Install Python 3.7 from Anaconda
- Open an Anaconda Prompt Terminal and create a new environment by running:

```
conda create --name hyperspy_env python=3.7
```

- Activate the above environment by:

```
WINDOWS: activate hyperspy_env
LINUX, macOS: source activate hyperspy_env
```

- Install the packages by running the following commands:

```
conda install hyperspy -c conda-forge
jupyter labextension install @jupyter-widgets/jupyterlab-manager
conda install -c conda-forge pyxem
conda install -c conda-forge atomap
```

- Not straightforward for beginner
- Usually not valid anymore after some time

HyperSpy bundle installer

- For beginner (and advanced) users
- Expected to work out of the box
- Very similar to Anaconda/Miniconda but includes relevant libraries

SuperSTEM DARESBURY

What makes HyperSpy today?

The University of Manchester

- Use tools and development practises which have proven to be successful for open source project
- HyperSpy is supported by its own community
- HyperSpy doesn’t rely on a specific research group/institution
- Peer-reviewed and open-source development
- HyperSpy is a mature library
  - API fairly stable
  - doesn’t break as much as before
- HyperSpy can be integrated easily in other software
  - Framework for the development of other libraries
  - HyperSpy will be split in the near future
    - Better modularity
    - Easier, faster implementation of new features

SuperSTEM DARESBURY

Achievement of HyperSpy and its community

The University of Manchester

- HyperSpy managed to built a *distributed* community of users/contributor
  - Led by its own contributors
  - Decision based on contributors consensus
- Motivate (at least not discourage) users
  - Pay attention to users feedback
  - From user to contributor: make the learning curve easier
- What are contributors doing?
  - Contribution to user guide, tutorials and online discussion
  - Code writing and/or code review

This is one way to make a library useful and sustainable

SuperSTEM DARESBURY

As individual is it worth contributing to HyperSpy?

The University of Manchester

- HyperSpy acknowledgement through zenodo DOI
  - One DOI for each release (important for reproducibility)
  - New contributors will be acknowledged
- Very good training
- Contribution to HyperSpy (or any other library) can be useful for career development
  - Github profile can be used as linkedin, etc.
  - Recognition by the community

SuperSTEM DARESBURY

As a group/organisation/company, why supporting HyperSpy?

The University of Manchester

- As PI/group leader
  - PI are not acknowledged but they can benefit a lot from the expertise gained through HyperSpy
  - Fairly useful “training” for post-doc/student
  - Remove dependence on expansive license for software or plugins
- Companies, conflict of interest?!
  - Some companies start to show interest in open source
  - Still a bit difficult: paradigm shift required
  - Customers are pushing enough to convince companies
  - Extend data processing capabilities of their software
    - By being compatible with open source software, companies can offer solution there could not afford otherwise